

ALMOST OPTIMAL SEQUENTIAL TESTS OF DISCRETE COMPOSITE HYPOTHESES

Georgios Fellouris and Alexander G. Tartakovsky

The University of Southern California

Abstract: We consider the problem of sequentially testing a simple null hypothesis, H_0 , versus a composite alternative hypothesis, H_1 , that consists of a finite set of densities. We study sequential tests that are based on thresholding of mixture-based likelihood ratio statistics and weighted generalized likelihood ratio statistics. It is shown that both sequential tests have several asymptotic optimality properties as error probabilities go to zero. First, for any weights, they minimize the expected sample size within a constant term under every scenario in H_1 and at least to first order under H_0 . Second, for appropriate weights that are specified up to a prior distribution, they minimize within an asymptotically negligible term a weighted expected sample size in H_1 . Third, for a particular prior distribution, they are almost minimax with respect to the expected Kullback–Leibler divergence until stopping. Furthermore, based on high-order asymptotic expansions for the operating characteristics, we propose prior distributions that lead to a robust behavior. Finally, based on asymptotic analysis as well as on simulation experiments, we argue that both tests have the same performance when they are designed with the same weights.

Key words and phrases: Asymptotic optimality, Generalized likelihood ratio, Minimax sequential tests, Mixture-based tests.

1. Introduction

Let $\{X_t\}_{t \in \mathbb{N}}$ be a sequence of independent and identically distributed (i.i.d.) random vectors with values in \mathbb{R}^d , $d \in \mathbb{N} = \{1, 2, \dots\}$, and common density f with respect to some non-degenerate, σ -finite measure $\nu(dx)$. We consider the problem of *sequentially* testing $H_0 : f \in \mathcal{A}_0$ versus $H_1 : f \in \mathcal{A}_1$, where \mathcal{A}_0 and \mathcal{A}_1 are two disjoint sets of densities with common support. That is, we assume that observations are acquired in a sequential manner and the goal is to select the correct hypothesis as soon as possible.

If $\{\mathcal{F}_t\}$ is the observed filtration, i.e., $\mathcal{F}_t = \sigma(X_1, \dots, X_t)$, a *sequential test* $\delta = (T, d_T)$ is a pair that consists of an $\{\mathcal{F}_t\}$ -stopping time, T , and an \mathcal{F}_T -measurable (terminal) decision rule, $d_T = d_T(X_1, \dots, X_T) \in \{0, 1\}$, that specifies which hypothesis is to be accepted once observations have stopped. In particular, H_j is accepted if $d_T = j$, i.e., $\{d_T = j\} = \{T < \infty, \delta \text{ accepts } H_j\}$, $j = 0, 1$.

An ideal sequential test should have the smallest possible expected sample size under both H_0 and H_1 , while controlling its error probabilities below given tolerance levels. Thus, if P_f is the underlying probability measure when X_1 has density f and E_f is the corresponding expectation, we will say that $\delta^o = (T^o, d_{T^o}^o) \in \mathcal{C}_{\alpha, \beta}$ is an *optimal* sequential test if

$$E_f[T^o] = \inf_{\delta \in \mathcal{C}_{\alpha, \beta}} E_f[T] \quad \forall f \in \mathcal{A}_0 \cup \mathcal{A}_1,$$

where $\mathcal{C}_{\alpha, \beta}$ is the class of sequential tests whose maximal type-I and type-II error probabilities are bounded above by α and β respectively, i.e.,

$$\mathcal{C}_{\alpha, \beta} = \left\{ \delta : \sup_{f \in \mathcal{A}_0} P_f(d_T = 1) \leq \alpha \quad \text{and} \quad \sup_{f \in \mathcal{A}_1} P_f(d_T = 0) \leq \beta \right\}.$$

[Wald and Wolfowitz \(1948\)](#) proved that an optimal sequential test exists when both hypotheses are simple, i.e., $\mathcal{A}_0 = \{f_0\}$ and $\mathcal{A}_1 = \{f_1\}$, and is given by the Sequential Probability Ratio Test (SPRT) that was proposed by [Wald \(1944\)](#) in his seminal work on Sequential Analysis:

$$S = \inf\{t \in \mathbb{N} : \Lambda_t^1 \notin (A^{-1}, B)\}, \quad d_S = \mathbb{1}_{\{\Lambda_S^1 \geq B\}}, \quad (1.1)$$

where $A, B > 1$ are constant thresholds selected so that $P_0(d_S = 1) = \alpha$ and $P_1(d_S = 0) = \beta$ and $\{\Lambda_t^1\}$ is the likelihood ratio statistic

$$\Lambda_t^1 = \prod_{n=1}^t \frac{f_1(X_n)}{f_0(X_n)}, \quad t \in \mathbb{N}. \quad (1.2)$$

In the case of composite hypotheses, it has only been possible to find sequential tests that are optimal in an asymptotic sense. More specifically, we will say that $\delta^0 \in \mathcal{C}_{\alpha, \beta}$ is *uniformly (first-order) asymptotically optimal*, if

$$E_f[T^0] = \inf_{\delta \in \mathcal{C}_{\alpha, \beta}} E_f[T] (1 + o(1)) \quad \forall f \in \mathcal{A}_0 \cup \mathcal{A}_1,$$

as $\alpha, \beta \rightarrow 0$. When, in particular, \mathcal{A}_0 and \mathcal{A}_1 can be embedded in an exponential family $\{f_\theta, \theta \in \Theta\}$ and Θ_1 is a subset of the natural parameter space Θ so that $\theta_0 \notin \Theta_1$ and

$$\mathcal{A}_0 = \{f_{\theta_0}\} \quad \text{and} \quad \mathcal{A}_1 = \{f_\theta, \theta \in \Theta_1\}, \quad (1.3)$$

it is well known (see, for example, [Lorden \(1973\)](#), [Pollak and Siegmund \(1975\)](#)) that the sequential test (1.1) is uniformly asymptotically optimal if Λ_t^1 is replaced either by the generalized likelihood-ratio (GLR) statistic, $\sup_{\theta \in \Theta} \Lambda_t^\theta$, or by a mixture-based likelihood ratio statistic, $\int_{\Theta} \Lambda_t^\theta w(\theta) d\theta$, where $w(\cdot)$ is some probability density function on Θ (weight function) and Λ_t^θ is defined as in (1.2) with f_1 replaced by f_θ . However, apart from certain tractable cases, both these statistics are not in general recursive and, as a result, they cannot be easily implemented on-line. Moreover, their computation at each step may be approximate, since it often requires discretization of the parameter space. These problems can be overcome if one uses the adaptive likelihood-ratio statistic, $\Lambda_t = \Lambda_{t-1}(f_{\theta_t^*}(X_t)/f_0(X_t))$, where θ_t^* is an estimator of θ that depends on the first $t - 1$ observations. However, this approach, initially developed by [Robbins and Siegmund \(1970, 1974\)](#) for power one tests and later extended by [Pavlov \(1990\)](#) and [Dragalin and Novikov \(1999\)](#) for multihypothesis sequential tests, generally leads to less efficient sequential tests, since one-stage delayed estimators use less information than the global MLE that is employed by the GLR statistic. Sequential testing of composite hypotheses in a Bayesian formulation with a small cost of observations was considered by [Chernoff \(1972\)](#); [Kiefer and Sacks \(1963\)](#); [Lai \(1988\)](#); [Lorden \(1967\)](#); [Schwarz \(1962\)](#) among others.

In the present paper, we consider the problem of sequential testing a simple null hypothesis against a *discrete* alternative consisting of a finite set of densities, i.e., we assume that

$$\mathcal{A}_0 = \{f_0\} \quad \text{and} \quad \mathcal{A}_1 = \{f_1, \dots, f_K\}, \quad (1.4)$$

where K is a positive integer. This hypothesis testing problem has two main motivations. First, it serves as an approximation to the continuous-parameter testing problem (1.3), in which Θ_1 is replaced by a finite subset $\{\theta_1, \dots, \theta_K\}$ of Θ_1 so that $f_j = f_{\theta_j}$, $j = 0, 1, \dots, K$. Indeed, as we mentioned above, the GLR statistic and mixture-based likelihood ratio statistics cannot always be easily implemented on-line and their computation may require discretization of the parameter space. With (1.4), we discretize the alternative hypothesis itself. This implies a loss of efficiency under P_θ when $\theta \notin \{\theta_1, \dots, \theta_K\}$, but it leads to sequential tests that are easily implementable on-line, a very important advantage for many applications.

Second, problem (1.4) naturally applies to multisample (also known as *multichannel* or *multisensor*) slippage problems, which have a wide range of applications (see,

e.g., Chernoff (1972); Tartakovsky *et al.* (2003, 2006)). As an example, consider the setup in which K sensors monitor different areas, a signal may be present in at most one of these areas and the goal is to detect signal presence without identifying its location. If additionally the sensors are statistically independent and sensor i takes i.i.d. observations $\{X_t^i\}_{t \in \mathbb{N}}$ with density g_1^i (resp. g_0^i) when signal is present (resp. absent), this problem turns out to be a special case of (1.4) with $X_t = (X_t^1, \dots, X_t^K)$ and

$$f_0(X_t) = \prod_{j=1}^K g_0^j(X_t^j), \quad f_i(X_t) = g_1^i(X_t^i) \prod_{\substack{j=1 \\ j \neq i}}^K g_0^j(X_t^j), \quad 1 \leq i \leq K. \quad (1.5)$$

For problem (1.4), we consider two sequential tests which are both parametrized by two vectors with positive components (*weights*), $\mathbf{q}_j = (q_j^1, \dots, q_j^K)$, $j = 0, 1$ and they both have the following structure: “stop the first time t at which either $\bar{\Lambda}_t \geq B$ or $\underline{\Lambda}_t \leq A^{-1}$ and select H_1 in the first case and H_0 in the latter”, where $\{\bar{\Lambda}_t\}$ and $\{\underline{\Lambda}_t\}$ are appropriate $\{\mathcal{F}_t\}$ -adapted statistics. For the first test, which we call *Mixture Likelihood Ratio Test* (MiLRT), the corresponding statistics are given by

$$\bar{\Lambda}_t = \sum_{i=1}^K q_1^i \Lambda_t^i \quad \text{and} \quad \underline{\Lambda}_t = \sum_{i=1}^K q_0^i \Lambda_t^i;$$

for the second test, which we call *Weighted Generalized Likelihood Ratio Test* (WGLRT), they are given by

$$\bar{\Lambda}_t = \max_{1 \leq i \leq K} (q_1^i \Lambda_t^i) \quad \text{and} \quad \underline{\Lambda}_t = \max_{1 \leq i \leq K} (q_0^i \Lambda_t^i),$$

where Λ_t^i is the likelihood ratio defined in (1.2) with f_1 replaced by f_i .

Tartakovsky *et al.* (2003) studied the GLRT, i.e., the WGLRT with *uniform* weights, $q_0^i = q_1^i = 1$, $1 \leq i \leq K$, in the multichannel setup (1.5) and established its asymptotic optimality. More specifically, it was shown that the GLRT is *second-order* asymptotically optimal, in the sense that it attains $\inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \mathbb{E}_i[T]$ within an $O(1)$ term for every $1 \leq i \leq K$, where $O(1)$ is asymptotically bounded as $\alpha, \beta \rightarrow 0$. Moreover, it was shown that, in the special case of completely asymmetric channels, the GLRT also attains $\inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \mathbb{E}_0[T]$ within an $O(1)$ term. (Here and in what follows we denote by \mathbb{P}_j the underlying probability measure when X_1 has density f_j and by \mathbb{E}_j the corresponding expectation, $j = 0, 1, \dots, K$.)

The first contribution of the present work is that this uniform, second-order asymptotic optimality property is established for both the MiLRT and the WGLRT with

arbitrary weights \mathbf{q}_0 and \mathbf{q}_1 in the more general setup of problem (1.4). However, the main question we want to answer is how to select these weights in order to obtain further “benefits”. In this direction, we show that if $\mathbf{p} = (p_1, \dots, p_K)$ is an arbitrary probability mass function, which can be interpreted as a prior distribution on \mathcal{H}_1 , and $\mathbf{q}_0, \mathbf{q}_1$ are selected so that

$$q_0^i = p_i \mathcal{L}_i \quad \text{and} \quad q_1^i = p_i / \mathcal{L}_i, \quad 1 \leq i \leq K, \quad (1.6)$$

then both tests attain $\inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \mathbb{E}^{\mathbf{P}}[T]$ within an $o(1)$ term, where $\mathbb{E}^{\mathbf{P}}$ is expectation with respect to the weighted probability measure $\mathbf{P}^{\mathbf{P}} = \sum_{i=1}^K p_i \mathbf{P}_i$ and the \mathcal{L} -numbers $\{\mathcal{L}_i\}$, formally introduced in (2.1), provide overshoot corrections that allow us to achieve this refined asymptotic optimality property.

In addition, we find a prior distribution $\hat{\mathbf{p}}$ which makes both tests *almost minimax* with respect to the expected Kullback–Leibler (KL) information (divergence) that is accumulated until stopping, in the sense that they attain within an $o(1)$ term

$$\inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \max_{1 \leq i \leq K} (I_i \mathbb{E}_i[T]),$$

where I_i is the KL-information number (see (2.2)). In this way, we generalize the corresponding result in [Fellouris and Tartakovsky \(2012\)](#), where this minimax problem was considered in the context of open-ended, mixture-based sequential tests.

Moreover, we compare numerically the tests with this (almost) least favorable prior distribution with some alternative choices for \mathbf{p} in the context of the multichannel problem (1.5) with channels that take exponential or Gaussian observations. Based on high-order asymptotic expansions for the operating characteristics of both tests, we find that selecting p_i to be proportional to I_i or \mathcal{L}_i leads to a much more robust behavior than the one induced by $\hat{\mathbf{p}}$, especially when the channels have very different signal strengths. Finally, based on these asymptotic expansions as well as on Monte Carlo simulations, we argue that both the WGLRT and the MiLRT have essentially the same performance when they are designed with the same weights.

The remainder of the paper is organized as follows. In [Section 2](#), we introduce basic notation and present some preliminary results. In [Section 3](#), we obtain asymptotic approximations to the operating characteristics of the two tests, whereas in [Section 4](#) we establish their asymptotic optimality properties. In [Section 5](#), we compare different specifications for \mathbf{p} and in [Section 6](#) we compare the tests using Monte Carlo simulations. We conclude in [Section 7](#).

2. Notation, Assumptions and Definitions

2.1. Elements of renewal theory

For every $1 \leq i \leq K$, we set $Z_t^i = \log \Lambda_t^i$, where Λ_t^i is given by (1.2) with f_1 replaced by f_i . We quantify the “distance” between f_i and f_0 using the \mathcal{L} -number

$$\mathcal{L}_i = \exp \left\{ - \sum_{n=1}^{\infty} n^{-1} \left[\mathbf{P}_0(Z_n^i > 0) + \mathbf{P}_i(Z_n^i \leq 0) \right] \right\}, \quad (2.1)$$

as well as the KL information numbers

$$I_i = \mathbf{E}_i[Z_1^i] = \int \log \left(\frac{f_i(x)}{f_0(x)} \right) f_1(x) \nu(dx), \quad (2.2)$$

$$I_0^i = \mathbf{E}_0[-Z_1^i] = \int \log \left(\frac{f_0(x)}{f_i(x)} \right) f_0(x) \nu(dx). \quad (2.3)$$

Without loss of generality, we assume that f_1, \dots, f_K are ordered with respect to their KL divergence from f_0 so that

$$I_0 = \min_{1 \leq i \leq K} I_0^i = I_0^1 = \dots = I_0^r < I_0^{r+1} \leq \dots \leq I_0^K. \quad (2.4)$$

Note that $r = 1$ corresponds to the asymmetric situation in which I_0 is attained by a unique index $i = 1$. On the other hand, $r = K$ corresponds to the completely symmetric situation in which I_0^i is the same for every $1 \leq i \leq K$. The latter case occurs, for example, in the multisample slippage problem (1.5) when $g_0^i = g_0$ and $g_1^i = g_1$, $1 \leq i \leq K$, i.e., when the densities do not depend on the population (or sensor, in a multisensor context).

In order to avoid trivial cases, we assume that f_i and f_0 do not coincide almost everywhere, which implies that $I_i, I_0^i > 0$ for every $1 \leq i \leq K$. We also assume throughout the paper that Z_1^i is non-arithmetic under \mathbf{P}_0 and \mathbf{P}_i and that $I_i, I_0^i < \infty$ for every $1 \leq i \leq K$. Then, if we define the first hitting times

$$\tau_c^i = \inf\{t : Z_t^i \geq c\}, \quad \sigma_c^i = \inf\{t : Z_t^i \leq -c\}, \quad c > 0,$$

it is well known that the overshoots $Z_{\tau_c^i}^i - c$ and $|Z_{\sigma_c^i}^i + c|$ have well defined asymptotic distributions under \mathbf{P}_i and \mathbf{P}_0 respectively, i.e.,

$$\mathcal{H}_i(x) = \lim_{c \rightarrow \infty} \mathbf{P}_i(Z_{\tau_c^i}^i - c \leq x), \quad \mathcal{H}_0^i(x) = \lim_{c \rightarrow \infty} \mathbf{P}_0(|Z_{\sigma_c^i}^i + c| \leq x), \quad x > 0,$$

and consequently, we can define the following Laplace transforms

$$\gamma_i = \int_0^\infty e^{-x} \mathcal{H}_i(dx), \quad \gamma_0^i = \int_0^\infty e^{-x} \mathcal{H}_0^i(dx),$$

which connect the KL-numbers with the \mathcal{L} -numbers as follows: $\mathcal{L}_i = \gamma_i I_i = \gamma_0^i I_0^i$ (see, e.g., Theorem 5 in [Lorden \(1977\)](#)). These quantities are very important, since they allow us to achieve with great accuracy the desired error probabilities of the SPRT, $\delta^i = (S^i, d_{S^i})$, for testing f_0 against f_i (that is, δ^i is given by (1.1) with Λ_t^1 replaced by Λ_t^i). Specifically, if $A = \gamma_0^i/\beta$ and $B = \gamma_i/\alpha$, then $P_0(d_{S^i} = 1) = \alpha(1 + o(1))$ and $P_i(d_{S^i} = 0) = \beta(1 + o(1))$ as $\alpha, \beta \rightarrow 0$ (see [Siegmund \(1975\)](#)).

If additionally second moments are finite, $E_i[(Z_1^i)^2], E_0[(Z_1^i)^2] < \infty$, then \mathcal{H}_i and \mathcal{H}_0^i have finite means (average limiting overshoots),

$$\kappa_i = \int_0^\infty x \mathcal{H}_i(dx), \quad \kappa_0^i = \int_0^\infty x \mathcal{H}_0^i(dx),$$

and we have the following asymptotic approximations for the expected sample sizes of the SPRT δ^i as $\alpha, \beta \rightarrow 0$ so that $\alpha|\log \beta| + \beta|\log \alpha| \rightarrow 0$:

$$E_i[S^i] = \frac{1}{I_i} (|\log \alpha| + \kappa_i + \log \gamma_i) + o(1), \quad (2.5)$$

$$E_0[S^i] = \frac{1}{I_0^i} (|\log \beta| + \kappa_0^i + \log \gamma_0^i) + o(1). \quad (2.6)$$

2.2. MiLRT and WGLRT

We will say that $\mathbf{q} = (q^1, \dots, q^K)$ is a *weight*, if $q_i > 0 \forall 1 \leq i \leq K$. For any weight \mathbf{q} , we set $|\mathbf{q}| = \sum_{i=1}^K q^i$ and we define

$$\Lambda_t(\mathbf{q}) = \sum_{i=1}^K q^i \Lambda_t^i, \quad \hat{\Lambda}_t(\mathbf{q}) = \max_{1 \leq i \leq K} \{q^i \Lambda_t^i\}, \quad (2.7)$$

$$Z_t(\mathbf{q}) = \log \Lambda_t(\mathbf{q}), \quad \hat{Z}_t(\mathbf{q}) = \log \hat{\Lambda}_t(\mathbf{q}). \quad (2.8)$$

The emphasis of this paper is on the MiLRT, $\delta_{\text{mi}} = (M, d_M)$, and the WGLRT, $\delta_{\text{gl}} = (N, d_N)$, which are parametrized by two arbitrary weights $\mathbf{q}_0, \mathbf{q}_1$ and are defined as follows:

$$M = \inf\{t : \Lambda_t(\mathbf{q}_1) \geq B \text{ or } \Lambda_t(\mathbf{q}_0) \leq A^{-1}\}, \quad d_M = \mathbb{1}_{\{\Lambda_M(\mathbf{q}_1) \geq B\}},$$

$$N = \inf\{t : \hat{\Lambda}_t(\mathbf{q}_1) \geq B \text{ or } \hat{\Lambda}_t(\mathbf{q}_0) \leq A^{-1}\}, \quad d_N = \mathbb{1}_{\{\hat{\Lambda}_N(\mathbf{q}_1) \geq B\}}.$$

Alternatively, if we introduce the following one-sided stopping times

$$M_B^1 = \inf\{t : \Lambda_t(\mathbf{q}_1) \geq B\}, \quad M_A^0 = \inf\{t : \Lambda_t(\mathbf{q}_0) \leq A^{-1}\},$$

$$N_B^1 = \inf\{t : \hat{\Lambda}_t(\mathbf{q}_1) \geq B\}, \quad N_A^0 = \inf\{t : \hat{\Lambda}_t(\mathbf{q}_0) \leq A^{-1}\},$$

δ_{mi} and δ_{gl} can be defined as follows

$$M = \min\{M_A^0, M_B^1\}, \quad d_M = \mathbb{1}_{\{M_B^1 \leq M_A^0\}}, \quad (2.9)$$

$$N = \min\{N_A^0, N_B^1\}, \quad d_N = \mathbb{1}_{\{N_B^1 \leq N_A^0\}}. \quad (2.10)$$

We also define the associated overshoots

$$\eta = [Z_M(\mathbf{q}_1) - \log B] \mathbb{1}_{\{d_M=1\}} - [Z_M(\mathbf{q}_0) + \log A] \mathbb{1}_{\{d_M=0\}}, \quad (2.11)$$

$$\hat{\eta} = [\hat{Z}_N(\mathbf{q}_1) - \log B] \mathbb{1}_{\{d_N=1\}} - [\hat{Z}_N(\mathbf{q}_0) + \log A] \mathbb{1}_{\{d_N=0\}}, \quad (2.12)$$

which play an important role in the asymptotic analysis of the operating characteristics of the two tests.

3. Asymptotic Approximations for the Operating Characteristics

In this section, we obtain asymptotic inequalities and approximations for the error probabilities and expected sample sizes of the MiLRT and the WGLRT. In order to do so, we rely on the following decompositions for $Z(\mathbf{q})$ and $\hat{Z}(\mathbf{q})$, which hold for every $1 \leq i \leq K$ and any weight $\mathbf{q} = (q^1, \dots, q^K)$,

$$Z_t(\mathbf{q}) = Z_t^i + \log q^i + Y_t^i(\mathbf{q}), \quad t \in \mathbb{N}, \quad (3.1)$$

$$\hat{Z}_t(\mathbf{q}) = Z_t^i + \log q^i + \hat{Y}_t^i(\mathbf{q}), \quad t \in \mathbb{N}, \quad (3.2)$$

where the sequences $Y^i(\mathbf{q})$ and $\hat{Y}^i(\mathbf{q})$ are defined as follows:

$$Y_t^i(\mathbf{q}) = \log \left(1 + \sum_{\substack{j=1 \\ j \neq i}}^K \frac{q^j}{q^i} \frac{\Lambda_t^j}{\Lambda_t^i} \right), \quad t \in \mathbb{N}, \quad (3.3)$$

$$\hat{Y}_t^i(\mathbf{q}) = \log \left(\max \left\{ 1, \max_{1 \leq j \neq i \leq K} \frac{q^j}{q^i} \frac{\Lambda_t^j}{\Lambda_t^i} \right\} \right), \quad t \in \mathbb{N}. \quad (3.4)$$

From the Strong Law of Large Numbers (SLLN) it follows that, for every $j \neq i$, $\mathbb{P}_i(\Lambda_t^j / \Lambda_t^i \rightarrow 0) = 1$. This implies that $Y^i(\mathbf{q})$ and $\hat{Y}^i(\mathbf{q})$ also converge to 0 \mathbb{P}_i -a.s., and consequently, they are *slowly changing* under \mathbb{P}_i (for a precise definition of “slowly changing” we refer to [Siegmund \(1985\)](#), page 190). Since Z_t^i is a random walk under \mathbb{P}_i , from this observation and decompositions (3.1)–(3.2) it follows that $Z(\mathbf{q})$ and $\hat{Z}(\mathbf{q})$ are *perturbed* random walks under \mathbb{P}_i .

Similarly, the SLLN implies that, *in the special case where $r = 1$* , $\mathbb{P}_0(\Lambda_t^j / \Lambda_t^1 \rightarrow 0) = 1$ for every $j > 1$. Therefore, $Y^1(\mathbf{q})$ and $\hat{Y}^1(\mathbf{q})$ also converge to 0 \mathbb{P}_0 -a.s. and

from (3.1)–(3.2) with $i = 1$ it follows that $Z(\mathbf{q})$ and $\hat{Z}(\mathbf{q})$ are perturbed random walks under P_0 when $r = 1$.

These properties allow us to apply nonlinear renewal theory for perturbed random walks (see Woodroffe (1976, 1982), Lai and Siegmund (1977, 1979), Siegmund (1985)) in order to obtain asymptotic approximations for the expected sample sizes of the tests δ_{mi} and δ_{gl} under P_i for every $1 \leq i \leq K$, as well as under P_0 when $r = 1$. An asymptotic approximation for $E_0[N]$ when $r > 1$ can be obtained based on the nonlinear renewal theory of Zhang (1988) using the following representation for N_A^0 :

$$N_A^0 = \inf \left\{ t : \ell_t^0 \geq \log A + \max_{1 \leq i \leq K} (\log q_0^i + \ell_t^i) \right\}, \quad (3.5)$$

where ℓ^j is the log-likelihood process under P_j for $j = 0, 1, \dots, K$, i.e.,

$$\ell_t^j = \sum_{n=1}^t \log f_j(X_n), \quad t \in \mathbb{N}. \quad (3.6)$$

For the latter approximation we also need some additional notation. Specifically, for any $1 \leq i \leq K$, we set $\mu_i = E_0[\log f_i(X_1)]$, so that $I_0^i = E_0[\log f_0(X_1)] - \mu_i$. Moreover, we set $\mu = \max_{1 \leq i \leq K} \mu_i$, so that $I_0 = E_0[\log f_0(X_1)] - \mu$, we define the r -dimensional random vector

$$W = (\log f_1(X_1) - \mu, \dots, \log f_r(X_1) - \mu), \quad (3.7)$$

and we denote by Σ its covariance matrix under P_0 . Finally, we set

$$d_r = \frac{h_r}{2\sqrt{I_0}}, \quad h_r = \int_{\mathbb{R}^r} \left(\max_{1 \leq i \leq r} x_i \right) \phi_\Sigma(x) dx, \quad (3.8)$$

where ϕ_Σ is the density of an r -dimensional, zero-mean, Gaussian random vector with covariance matrix Σ .

3.1. Asymptotic bounds for the error probabilities

We start with the following lemma.

Lemma 1. *For any $1 \leq i \leq K$,*

$$E_i[e^{-\eta} \mathbb{1}_{\{d_M=1\}}] \rightarrow \gamma_i, \quad E_i[e^{-\hat{\eta}} \mathbb{1}_{\{d_N=1\}}] \rightarrow \gamma_i \quad \text{as } A, B \rightarrow \infty. \quad (3.9)$$

If additionally $r = 1$, then

$$E_0[e^{-\eta} \mathbb{1}_{\{d_M=0\}}] \rightarrow \gamma_0^1, \quad E_0[e^{-\hat{\eta}} \mathbb{1}_{\{d_N=0\}}] \rightarrow \gamma_0^1 \quad \text{as } A, B \rightarrow \infty. \quad (3.10)$$

Proof. We will only prove the first assertions in (3.9) and (3.10), since the other ones can be proven in an identical way.

Since $M = M_B^1 = \inf\{t : Z_t(\mathbf{q}_1) \geq \log B\}$ and $\eta = Z_{M_B^1}(\mathbf{q}_1) - \log B$ on $\{d_M = 1\} = \{M_B^1 \leq M_A^0\}$, and $\{Z_t(\mathbf{q}_1) = Z_t^i + \log q_1^i + Y_t^i(\mathbf{q}_1)\}$ is a perturbed random walk under P_i , from nonlinear renewal theory (see, e.g., Theorem 9.12 in Siegmund (1985)) it follows that η converges in distribution to \mathcal{H}_i under P_i on $\{d_M = 1\}$. Therefore, the Bounded Convergence Theorem yields $E_i[e^{-\eta} \mathbb{1}_{\{d_M=1\}}] \rightarrow \gamma_i$.

Since $M = M_A^0 = \inf\{t : -Z_t(\mathbf{q}_0) \geq \log A\}$ and $\eta = |Z_{M_A^0}(\mathbf{q}_0) + \log A|$ on $\{d_M = 0\} = \{M_B^1 > M_A^0\}$, and $\{-Z_t(\mathbf{q}_0) = -Z_t^1 - \log q_0^1 - Y_t^1(\mathbf{q}_0)\}$ is a perturbed random walk under P_0 when $r = 1$, the same argument as above applies to show that $E_0[e^{-\eta} \mathbb{1}_{\{d_M=0\}}] \rightarrow \gamma_0^1$. \square

The following theorem provides exact and asymptotic upper bounds on the error probabilities of δ_{mi} and δ_{gl} .

Theorem 1. (a) For any $A, B > 1$,

$$P_0(d_M = 1) \leq \frac{|\mathbf{q}_1|}{B}, \quad P_0(d_N = 1) \leq \frac{|\mathbf{q}_1|}{B}, \quad (3.11)$$

$$P_i(d_M = 0) \leq \frac{1}{A q_0^i}, \quad P_i(d_N = 0) \leq \frac{1}{A q_0^i}, \quad 1 \leq i \leq K. \quad (3.12)$$

(b) As $A, B \rightarrow \infty$,

$$P_0(d_M = 1) = \frac{1}{B} \left(\sum_{j=1}^K q_1^j \gamma_j \right) (1 + o(1)), \quad (3.13)$$

$$P_0(d_N = 1) \leq \frac{1}{B} \left(\sum_{j=1}^K q_1^j \gamma_j \right) (1 + o(1)). \quad (3.14)$$

If additionally $r = 1$, then for every $1 \leq i \leq K$

$$P_i(d_M = 0) \leq \frac{\gamma_0^1}{q_0^i A} (1 + o(1)), \quad P_i(d_N = 0) \leq \frac{\gamma_0^1}{q_0^i A} (1 + o(1)). \quad (3.15)$$

Proof. Let us define the probability measure $P^{\mathbf{q}_1} = \frac{1}{|\mathbf{q}_1|} \sum_{i=1}^K q_1^i P_i$ and denote by $E^{\mathbf{q}_1}$ expectation with respect to $P^{\mathbf{q}_1}$. Since

$$\frac{dP^{\mathbf{q}_1}}{dP_0} \Big|_{\mathcal{F}_t} = \frac{1}{|\mathbf{q}_1|} \sum_{i=1}^K q_1^i \Lambda_t^i = \frac{1}{|\mathbf{q}_1|} e^{Z_t(\mathbf{q}_1)},$$

changing the measure $P_0 \mapsto P^{\mathbf{q}_1}$ we have

$$P_0(d_M = 1) = |\mathbf{q}_1| \mathbb{E}^{\mathbf{q}_1}[e^{-Z_M(\mathbf{q}_1)} \mathbb{1}_{\{d_M=1\}}] = \sum_{i=1}^K q_1^i \mathbb{E}_i[e^{-Z_M(\mathbf{q}_1)} \mathbb{1}_{\{d_M=1\}}]. \quad (3.16)$$

Since $Z_M(\mathbf{q}_1) = \log B + \eta$ on $\{d_M = 1\}$, we obtain

$$P_0(d_M = 1) = \frac{1}{B} \sum_{i=1}^K q_1^i \mathbb{E}_i[e^{-\eta} \mathbb{1}_{\{d_M=1\}}]. \quad (3.17)$$

Since η is positive, the first inequality in (3.11) immediately follows from (3.17), whereas (3.13) follows from (3.9). A similar argument as the one that led to (3.16), along with the fact that $Z_t(\mathbf{q}_1) \geq \hat{Z}_t(\mathbf{q}_1)$, yields

$$\begin{aligned} P_0(d_N = 1) &= \sum_{i=1}^K q_1^i \mathbb{E}_i[e^{-Z_N(\mathbf{q}_1)} \mathbb{1}_{\{d_N=1\}}] \\ &\leq \sum_{i=1}^K q_1^i \mathbb{E}_i[e^{-\hat{Z}_N(\mathbf{q}_1)} \mathbb{1}_{\{d_N=1\}}] \leq \frac{1}{B} \sum_{i=1}^K q_1^i \mathbb{E}_i[e^{-\hat{\eta}} \mathbb{1}_{\{d_N=1\}}]. \end{aligned} \quad (3.18)$$

The last inequality and the fact that $\hat{\eta}$ is positive imply the second inequality in (3.11), whereas (3.14) follows from (3.9).

Finally, changing the measure $P_i \mapsto P_0$, we obtain

$$P_i(d_M = 0) = \mathbb{E}_0[e^{Z_M^i} \mathbb{1}_{\{d_M=0\}}]. \quad (3.19)$$

Since $Z_M^i = Z_M(\mathbf{q}_0) - \log q_0^i - Y_M^i(\mathbf{q}_0)$ (recall (3.1)), $Z_M(\mathbf{q}_0) = -\log A - \eta$ on $\{d_M = 0\}$ (recall (2.11)) and $Y_M^i(\mathbf{q}_0) \geq 0$, it follows that $Z_M^i \leq -\log(Aq_0^i) - \eta$ on $\{d_M = 0\}$ and, consequently, (3.19) becomes

$$P_i(d_M = 0) \leq \frac{1}{Aq_0^i} \mathbb{E}_0[e^{-\eta} \mathbb{1}_{\{d_M=0\}}].$$

Since η is positive, we obtain the first inequality in (3.12), whereas from (3.10) we obtain the first inequality in (3.15). The remaining inequalities in (3.12) and (3.15) can be shown in a similar way. \square

From Theorem 1(a) it is clear that when A, B are selected according to

$$A_\beta(\mathbf{q}_0) = \frac{1}{\beta \min_{1 \leq i \leq K} q_0^i}, \quad B_\alpha(\mathbf{q}_1) = \frac{|\mathbf{q}_1|}{\alpha}, \quad (3.20)$$

then $\delta_{\text{mi}}, \delta_{\text{gl}} \in \mathcal{C}_{\alpha, \beta}$. Moreover, from Theorem 1(b) it follows that we can obtain sharper inequalities if we correct for the overshoots selecting A, B as follows

$$A_{\beta}(\mathbf{q}_0) = \frac{\gamma_0^1}{\beta \min_{1 \leq i \leq K} q_0^i}, \quad B_{\alpha}(\mathbf{q}_1) = \frac{\sum_{j=1}^K q_1^j \gamma_j}{\alpha}. \quad (3.21)$$

Indeed, with this selection of the thresholds we have $P_0(d_M = 1) = \alpha(1 + o(1))$, $P_0(d_N = 1) \leq \alpha(1 + o(1))$ and if additionally $r = 1$, $\max_{1 \leq i \leq K} P_i(d_M = 0) \leq \beta(1 + o(1))$ and $\max_{1 \leq i \leq K} P_i(d_N = 0) \leq \beta(1 + o(1))$.

3.2. Asymptotic approximations to expected sample sizes

In order to obtain asymptotic approximations to the expected sample sizes of the MiLRT and the WGLRT, we will make the following assumptions, which will be needed for all the results in the rest of the paper:

$$(A1) \ E_i[(Z_1^i)^2] < \infty \quad \text{and} \quad E_0[(Z_1^i)^2] < \infty, \quad 1 \leq i \leq K;$$

$$(A2) \ \alpha, \beta \rightarrow 0 \text{ so that } |\log \alpha|/|\log \beta| \rightarrow k, \text{ where } k \in (0, \infty);$$

$$(A3) \ \text{For } T = M \text{ or } T = N, A \text{ and } B \text{ are selected so that as } \alpha, \beta \rightarrow 0$$

$$k_0 \alpha (1 + o(1)) \leq P_0(d_T = 1) \leq \alpha (1 + o(1)), \quad (3.22)$$

$$k_1 \beta (1 + o(1)) \leq \max_{1 \leq i \leq K} P_i(d_T = 0) \leq \beta (1 + o(1)), \quad (3.23)$$

or equivalently,

$$|\log \alpha| + o(1) \leq |\log P_0(d_T = 1)| \leq |\log \alpha| + |\log k_0| + o(1), \quad (3.24)$$

$$|\log \beta| + o(1) \leq |\log \max_{1 \leq i \leq K} P_i(d_T = 0)| \leq |\log \beta| + |\log k_1| + o(1), \quad (3.25)$$

where $k_0, k_1 \in (0, 1)$ are fixed constants, not necessarily the same for δ_{mi} and δ_{gl} .

The second moment conditions (A1) on the log-likelihood ratio Z_1^i are required even for the asymptotic approximations (2.5)–(2.6) to the performance of the SPRT for testing f_0 against f_i . Assumption (A2) concerns the relative rates with which α and β go to 0 and requires that α should not go to 0 exponentially faster than β and vice-versa. Note, however, that α can still be much smaller than β (or vice versa), a natural requirement in many applications. Assumption (A3) requires that the thresholds for both the MiLRT and the WGLRT are designed so that the probabilities of the type-I and type-II errors are *asymptotically* bounded by (and at the same time not much smaller than) α and β respectively. As the following lemma suggests, (A3) connects the thresholds

A and B with the desired error probabilities α and β , so that we do not need to impose additional (to (A2)) constraints to the relative rates with which A and B go to infinity.

Lemma 2. *If (A3) holds, then $\log B = |\log \alpha| + O(1)$ and $\log A = |\log \beta| + O(1)$.*

Proof. From (3.11) we know that $\log B \leq |\log P_0(d_M = 1)| + |\mathbf{q}_1|$, whereas from (A3), and in particular (3.24), it follows that $|\log P_0(d_M = 1)| \leq |\log \alpha| + |\log k_0| + o(1)$, which proves $\log B = |\log \alpha| + O(1)$. The second relationship can be shown in a similar way. \square

Theorem 2. *If conditions (A1)–(A3) hold, then*

(a) *for every $1 \leq i \leq K$,*

$$I_i \mathbb{E}_i[M] = \log B + \kappa_i - \log q_1^i + o(1), \quad (3.26)$$

$$I_i \mathbb{E}_i[N] = \log B + \kappa_i - \log q_1^i + o(1); \quad (3.27)$$

(b) *for $r = 1$,*

$$I_0 \mathbb{E}_0[M] = \log A + \kappa_0^1 + \log q_0^1 + o(1), \quad (3.28)$$

$$I_0 \mathbb{E}_0[N] = \log A + \kappa_0^1 + \log q_0^1 + o(1); \quad (3.29)$$

(c) *for $r > 1$,*

$$I_0 \mathbb{E}_0[M] = \log A + 2 d_r \sqrt{\log A} + O(1), \quad (3.30)$$

$$I_0 \mathbb{E}_0[N] = \log A + 2 d_r \sqrt{\log A} + O(1), \quad (3.31)$$

where d_r is defined in (3.8).

Proof. (a) Asymptotic approximations (3.26) and (3.27) can be relatively easily established using nonlinear renewal theory. Specifically, starting from representation (3.1) and applying the Nonlinear Renewal Theorem (see Theorem 9.28 in Siegmund (1985)), it can be shown (as in Theorem 2.1 of Fellouris and Tartakovsky (2012)) that $I_i \mathbb{E}_i[M_B^1]$ is equal to the right-hand side of (3.26) as $B \rightarrow \infty$. Therefore, to prove (3.26) it suffices to show that $\mathbb{E}_i[M_B^1 - M] = o(1)$ as $A, B \rightarrow \infty$, or equivalently as $\alpha, \beta \rightarrow 0$. To this end, note that

$$0 \leq M_B^1 - M = [M_B^1 - M_A^0] \mathbb{1}_{\{d_M=0\}} \leq M_B^1 \mathbb{1}_{\{d_M=0\}}.$$

Applying the Cauchy–Schwartz inequality, we obtain

$$\mathbb{E}_i[M_B^1 \mathbb{1}_{\{d_M=0\}}] \leq \sqrt{\mathbb{E}_i[(M_B^1)^2] \mathbb{P}_i(d_M = 0)}. \quad (3.32)$$

From (3.1) and (3.3) it is clear that $Z_t(\mathbf{q}_1) \geq Z_t^i + \log q_1^i$, $t \in \mathbb{N}$, thus,

$$M_B^1 \leq \inf\{t : Z_t^i \geq \log(B/q_1^i)\}.$$

Consequently, from Theorem 8.1 in Gut (2008) it follows that, since (A1) holds,

$$(I_i)^2 \mathbb{E}_i[(M_B^1)^2] \leq (\log(B/q_1^i))^2(1 + o(1)).$$

From the latter inequality and Lemma 2 we conclude that

$$\mathbb{E}_i[(M_B^1)^2] = O((\log B)^2) = O(|\log \alpha|^2).$$

Moreover, since (A3) implies $\mathbb{P}_i(d_M = 0) \leq \beta(1 + o(1))$, (3.32) becomes

$$\mathbb{E}_i[M_B^1 \mathbb{1}_{\{d_M=0\}}] = O(|\log \alpha|^2 \beta)$$

and from (A2) we conclude that the upper bound goes to 0. This completes the proof of (3.26), whereas the proof of (3.27) is analogous.

(b) From representation (3.1) and the Nonlinear Renewal Theorem it follows that $I_0 \mathbb{E}_0[M_A^0]$ is equal to the right-hand side of (3.28) as $A \rightarrow \infty$. Then, similarly to (a), we can show that $\mathbb{E}_0[M_A^0 - M] = o(1)$. The proof of (3.29) follows similar steps.

(c) In order to prove (3.31), we start from representation (3.5) and apply nonlinear renewal theory of Zhang (1988). As a result, it can be shown (analogously to Lemma 2.1 of Dragalin (1999)) that $I_0 \mathbb{E}_0[N_A^0]$ is equal to the right-hand side of (3.31). Thus, it suffices to show that $\mathbb{E}_0[N_A^0] = \mathbb{E}_0[N] + o(1)$, which can be done in just the same way as in (a) and (b). \square

Remark 1. Asymptotic approximation (3.31) can be further improved (up to the negligible term $o(1)$), if stronger integrability conditions are postulated on the vector W defined in (3.7). Specifically, if in addition we assume the third moment condition $\mathbb{E}_0[||W||^3] < \infty$ as well as the Cramer-type condition $\limsup_{||t|| \rightarrow \infty} \mathbb{E}_0[e^{j \langle t, W \rangle}] < 1$, where j is the imaginary unit, $t = (t_1, \dots, t_r)$ and $\langle t, W \rangle = \sum_{l=1}^r t_l W_l$, then the following expansion holds

$$\begin{aligned} I_0 \mathbb{E}_0[N] &= \log A + 2 d_r \sqrt{\log A + d_r^2} + \frac{h_r^2}{2I_0} + \kappa_0^1 \\ &\quad + \int_{\mathbb{R}^r} \left\{ \max_{1 \leq i \leq r} (x_i) [\mathcal{P}(x) + \lambda(\mathbf{q}_0) \Sigma^{-1} x'] \right\} \phi_{\Sigma}(x) dx + o(1), \end{aligned}$$

where $\lambda(\mathbf{q}_0) = (\log q_0^1, \dots, \log q_0^r)$ and \mathcal{P} is a third-degree polynomial whose coefficients depend on the P_0 -cumulants of W (see Bhattacharya and Rao (1986)). This

approximation can be derived similarly to Theorem 3.3 of [Dragalin et al. \(2000\)](#) based on nonlinear renewal theory of [Zhang \(1988\)](#).

Corollary 1. *Suppose that (A1)–(A3) hold with $k_0 = 1$, i.e., A and B are selected so that $P_0(d_M = 1) \sim \alpha$ and $P_0(d_N = 1) \sim \alpha$. Then,*

$$I_i E_i[M] = |\log \alpha| + \log \left(\sum_{j=1}^K q_1^j \gamma_j \right) + \kappa_i - \log q_1^i + o(1), \quad (3.33)$$

$$I_i E_i[N] \leq |\log \alpha| + \log \left(\sum_{j=1}^K q_1^j \gamma_j \right) + \kappa_i - \log q_1^i + o(1). \quad (3.34)$$

Proof. From (3.13)–(3.14) it follows that

$$\log B = |\log P_0(d_M = 1)| + \log \left(\sum_{j=1}^K q_1^j \gamma_j \right) + o(1),$$

$$\log B \leq |\log P_0(d_N = 1)| + \log \left(\sum_{j=1}^K q_1^j \gamma_j \right) + o(1).$$

Moreover, from (3.24) and the assumption that $k_0 = 1$ we have

$$|\log P_0(d_M = 1)| = |\log \alpha| + o(1) \quad \text{and} \quad |\log P_0(d_N = 1)| = |\log \alpha| + o(1).$$

From these two relationships and Theorem 2(a) we obtain the desired result. \square

4. Asymptotic Optimality Properties

In this section, we establish the asymptotic optimality properties of the MiLRT and the WGLRT.

4.1. Uniform asymptotic optimality

First, we show that both tests minimize the expected sample size within an $O(1)$ term (i.e., to second order) under every P_i , $1 \leq i \leq K$ and at least to first order under P_0 .

Theorem 3. *Suppose that conditions (A1)–(A3) hold and that A, B are selected so that $\delta_{\text{mi}}, \delta_{\text{gl}} \in \mathcal{C}_{\alpha, \beta}$.*

(a) *For every $1 \leq i \leq K$,*

$$E_i[M] = \inf_{\delta \in \mathcal{C}_{\alpha, \beta}} E_i[T] + O(1), \quad (4.1)$$

$$E_i[N] = \inf_{\delta \in \mathcal{C}_{\alpha, \beta}} E_i[T] + O(1). \quad (4.2)$$

(b) If $r = 1$, then

$$\mathbb{E}_0[M] = \inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \mathbb{E}_0[T] + O(1), \quad (4.3)$$

$$\mathbb{E}_0[N] = \inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \mathbb{E}_0[T] + O(1), \quad (4.4)$$

whereas if $r > 1$,

$$\mathbb{E}_0[M] = \inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \mathbb{E}_0[T] (1 + o(1)), \quad (4.5)$$

$$\mathbb{E}_0[N] = \inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \mathbb{E}_0[T] (1 + o(1)). \quad (4.6)$$

Proof. (a) From (2.5) it is clear that

$$I_i \inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \mathbb{E}_i[T] \geq |\log \alpha| + O(1), \quad (4.7)$$

whereas from Theorem 2(a) and Lemma 2 it follows that

$$I_i \mathbb{E}_i[M] = \log B + O(1) = |\log \alpha| + O(1),$$

which proves (4.1). The proof of (4.2) is similar.

(b) From (2.6) it is clear that for every $1 \leq i \leq K$ we have

$$\inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \mathbb{E}_0[T] \geq \frac{|\log \beta|}{I_0^i} + O(1), \quad (4.8)$$

thus, recalling from (2.4) that $I_0 = \min_{1 \leq i \leq K} I_0^i$, we obtain

$$\inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \mathbb{E}_0[T] \geq \frac{|\log \beta|}{I_0} + O(1). \quad (4.9)$$

But from Theorem 2(b) and Lemma 2 it follows that

$$I_0 \mathbb{E}_0[M] = \begin{cases} \log A + O(1) = |\log \beta| + O(1), & \text{if } r = 1, \\ \log A (1 + o(1)) = |\log \beta| (1 + o(1)), & \text{if } r > 1, \end{cases} \quad (4.10)$$

which implies (4.3) and (4.5). The proofs of (4.4) and (4.6) are similar. \square

4.2. Almost optimality

In what follows, we denote by $\delta_{\text{mi}}^*(\mathbf{p}) = (M^*(\mathbf{p}), d_{M^*(\mathbf{p})})$ and $\delta_{\text{gl}}^*(\mathbf{p}) = (N^*(\mathbf{p}), d_{N^*(\mathbf{p})})$ the MiLRT and the WGLRT with weights given by (1.6), i.e.

$$q_1^i = \frac{p_i}{\mathcal{L}_i} \quad \text{and} \quad q_0^i = p_i \mathcal{L}_i, \quad 1 \leq i \leq K, \quad (4.11)$$

where $\mathbf{p} = (p_1, \dots, p_K)$, $p_i > 0$ for every $1 \leq i \leq K$ and $\sum_{i=1}^K p_i = 1$. Our goal is to show that $\delta_{\text{mi}}^*(\mathbf{p})$ and $\delta_{\text{gl}}^*(\mathbf{p})$ attain $\inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \mathbb{E}^{\mathbf{p}}[T]$ asymptotically within an $o(1)$ term, where $\mathbb{E}^{\mathbf{p}}$ is expectation with respect to the weighted probability measure

$\mathbf{P}\mathbf{P} = \sum_{i=1}^K p_i \mathbf{P}_i$. Before doing so, note that Corollary 1 implies that if B is selected so that $\mathbf{P}_0(d_{M^*}(\mathbf{p}) = 1) \sim \alpha$ and $\mathbf{P}_0(d_{N^*}(\mathbf{p}) = 1) \sim \alpha$, then

$$\mathbb{E}_i[M^*(\mathbf{p})] = \frac{1}{I_i} \left[|\log \alpha| + \kappa_i + \log \gamma_i + C_i(\mathbf{p}) \right] + o(1), \quad (4.12)$$

$$\mathbb{E}_i[N^*(\mathbf{p})] \leq \frac{1}{I_i} \left[|\log \alpha| + \kappa_i + \log \gamma_i + C_i(\mathbf{p}) \right] + o(1), \quad (4.13)$$

where we have used the fact that $\mathcal{L}_i = \gamma_i I_i$, $1 \leq i \leq K$ and we have introduced the following notation

$$C_i(\mathbf{p}) = \log \left(\sum_{j=1}^K \frac{p_j}{I_j} \right) - \log \frac{p_i}{I_i}, \quad 1 \leq i \leq K. \quad (4.14)$$

Theorem 4. *Suppose that conditions (A1)–(A3) hold with $k = 1$, i.e., $\alpha, \beta \rightarrow 0$ so that $|\log \alpha| \sim |\log \beta|$. Then*

$$\inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \mathbb{E}\mathbf{P}[T] = \sum_{i=1}^K \frac{p_i}{I_i} \left[|\log \alpha| + \kappa_i + \log \gamma_i + C_i(\mathbf{p}) \right] + o(1). \quad (4.15)$$

Moreover, if A, B are selected so that $\delta_{\text{mi}}^*(\mathbf{p})$ and $\delta_{\text{gl}}^*(\mathbf{p})$ belong to $\mathcal{C}_{\alpha, \beta}$ and $k_0 = 1$, i.e., $\mathbf{P}_0(d_{M^*}(\mathbf{p}) = 1) \sim \alpha$ and $\mathbf{P}_0(d_{N^*}(\mathbf{p}) = 1) \sim \alpha$, then

$$\begin{aligned} \inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \mathbb{E}\mathbf{P}[T] &= \mathbb{E}\mathbf{P}[M^*(\mathbf{p})] + o(1), \\ \inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \mathbb{E}\mathbf{P}[T] &= \mathbb{E}\mathbf{P}[N^*(\mathbf{p})] + o(1). \end{aligned}$$

In order to prove this theorem, we formulate our sequential testing problem as a Bayesian sequential decision problem with $K + 1$ states, $\mathbf{H}_0 : f = f_0$ and $\mathbf{H}_1^i : f = f_i$, $1 \leq i \leq K$ and two possible actions upon stopping, either accepting \mathbf{H}_0 or $\mathbf{H}_1 = \cup_i \mathbf{H}_1^i$. Moreover, we denote by c the sampling cost per observation and by w_1 (resp. w_0) the loss associated with accepting \mathbf{H}_0 (resp. \mathbf{H}_1) when the correct hypothesis is \mathbf{H}_1 (resp. \mathbf{H}_0). We also define the probability measure $\mathbf{P}^\pi = \pi \mathbf{P}_0 + (1 - \pi) \mathbf{P}\mathbf{P}$, which means that $\pi = \mathbf{P}^\pi(\mathbf{H}_0)$ is the prior probability of \mathbf{H}_0 and $p_i = \mathbf{P}^\pi(\mathbf{H}_1^i | \mathbf{H}_1)$ is the prior probability of $f = f_i$ given that \mathbf{H}_1 is correct.

The integrated risk of a sequential test $\delta = (T, d_T)$ is defined as the sum $\mathcal{R}(\delta) = \mathcal{R}_c(T) + \mathcal{R}_s(d_T)$, where $\mathcal{R}_c(T)$ is the integrated risk due to sampling and $\mathcal{R}_s(d_T)$ is

the integrated risk due to a wrong decision upon stopping, i.e.,

$$\begin{aligned}\mathcal{R}_c(T) &= c \mathbf{E}^\pi[T] = c \left[\pi \mathbf{E}_0[T] + (1 - \pi) \mathbf{E}^{\mathbf{P}}[T] \right], \\ \mathcal{R}_s(d_T) &= \mathbf{E}^\pi[w_0 \mathbb{1}_{\{d_T=1\}} | \mathbf{H}_0] + \mathbf{E}^\pi[w_1 \mathbb{1}_{\{d_T=0\}} | \mathbf{H}_1] \\ &= \pi w_0 \mathbf{P}_0(d_T = 1) + (1 - \pi) w_1 \mathbf{P}^{\mathbf{P}}(d_T = 0).\end{aligned}$$

The Bayesian sequential decision problem is to find an optimal (*Bayes*) sequential test that attains the *Bayes risk*, $\mathcal{R}^* = \inf_\delta \mathcal{R}(\delta)$. It is well known that the solution to this problem does not have a simple structure (see, e.g., [Chow *et al.* \(1971\)](#)). However, from the seminal work of [Lorden \(1977\)](#) on finite-state sequential decision making it follows that $\delta_{\text{mi}}^*(\mathbf{p})$ and $\delta_{\text{gl}}^*(\mathbf{p})$ are *almost Bayes* when the thresholds A and B are chosen as

$$A_c = \frac{1 - \pi}{\pi} \frac{w_1}{c} \quad \text{and} \quad B_c = \frac{\pi}{1 - \pi} \frac{w_0}{c}. \quad (4.16)$$

More specifically, denote by $\delta_{\text{mi},c}^*(\mathbf{p}) = (M_c^*(\mathbf{p}), d_{M_c^*}(\mathbf{p}))$ and $\delta_{\text{gl},c}^*(\mathbf{p}) = (N_c^*(\mathbf{p}), d_{N_c^*}(\mathbf{p}))$ the sequential tests $\delta_{\text{mi}}^*(\mathbf{p})$ and $\delta_{\text{gl}}^*(\mathbf{p})$ when the thresholds are given by A_c and B_c . Under the integrability condition (A1), it follows from [Lorden \(1977\)](#) that

$$\mathcal{R}(\delta_{\text{mi},c}^*(\mathbf{p})) - \mathcal{R}^* = o(c) \quad \text{and} \quad \mathcal{R}(\delta_{\text{gl},c}^*(\mathbf{p})) - \mathcal{R}^* = o(c). \quad (4.17)$$

The proof of Theorem 4 relies on this third-order Bayesian asymptotic optimality property, which requires symmetric thresholds (4.16) and is the reason why we assumed in Theorem 4 that error probabilities go to 0 with the same rate.

Proof. In order to lighten the notation, we omit the dependence on the prior distribution \mathbf{p} and write simply $\delta_{\text{mi}}^* = (M^*, d_{M^*})$ and $\delta_{\text{mi},c}^* = (M_c^*, d_{M_c^*})$ instead of $\delta_{\text{mi}}^*(\mathbf{p}) = (M^*(\mathbf{p}), d_{M^*}(\mathbf{p}))$ and $\delta_{\text{mi},c}^*(\mathbf{p}) = (M_c^*(\mathbf{p}), d_{M_c^*}(\mathbf{p}))$ (and similarly for the WGLRT).

From Corollary 1 it is clear that the right-hand side in (4.15) is attained by δ_{mi}^* and δ_{gl}^* when their thresholds are selected so that $\mathbf{P}_0(d_{M^*} = 1) \sim \alpha$ and $\mathbf{P}_0(d_{N^*} = 1) \sim \alpha$. If additionally $\delta_{\text{mi}}^*, \delta_{\text{gl}}^* \in \mathcal{C}_{\alpha,\beta}$, then $\inf_{\delta \in \mathcal{C}_{\alpha,\beta}} \mathbf{E}^{\mathbf{P}}[T]$ is attained by these two tests to within an $o(1)$ term. Thus, it suffices to establish (4.15).

Consider the class of sequential tests

$$\mathcal{C}_{\alpha,\beta}^{\mathbf{P}} = \{\delta : \mathbf{P}_0(d_T = 1) \leq \alpha \quad \text{and} \quad \mathbf{P}^{\mathbf{P}}(d_T = 0) \leq \beta\}.$$

Since $\mathcal{C}_{\alpha,\beta} \subset \mathcal{C}_{\alpha,\beta}^{\mathbf{P}}$, we have $\inf_{\delta \in \mathcal{C}_{\alpha,\beta}} \mathbf{E}^{\mathbf{P}}[T] \geq \inf_{\delta \in \mathcal{C}_{\alpha,\beta}^{\mathbf{P}}} \mathbf{E}^{\mathbf{P}}[T]$. Thus, it suffices to show that

$$\inf_{\delta \in \mathcal{C}_{\alpha,\beta}^{\mathbf{P}}} \mathbf{E}^{\mathbf{P}}[T] = \sum_{i=1}^K \frac{p_i}{I_i} \left[|\log \alpha| + \kappa_i + \log \gamma_i + C_i(\mathbf{p}) \right] + o(1). \quad (4.18)$$

Consider now the sequential test $\delta_{\text{mi},c}^* = (M_c^*, d_{M_c^*})$ with thresholds A_c and B_c selected so that $P_0(d_{M_c^*} = 1) = \alpha$ and $PP(d_{M_c^*} = 0) = \beta$. From Corollary 1 it is clear that $EP[M_c^*]$ is equal to the right-hand side in (4.18) as $c \rightarrow 0$, which means that it suffices to show that

$$\inf_{\delta \in \mathcal{C}_{\alpha,\beta}^P} EP[T] = EP[M_c^*] + o(1),$$

where $o(1)$ is an asymptotically negligible term as $c \rightarrow 0$. More specifically, if δ is an arbitrary sequential test in $\mathcal{C}_{\alpha,\beta}^P$, we need to show that, for sufficiently small c , $|EP[T] - EP[M_c^*]|$ is bounded above by an arbitrarily small, but fixed number.

First of all, we observe that

$$\begin{aligned} \mathcal{R}_s(d_T) &= \pi w_0 P_0(d_T = 1) + (1 - \pi) w_1 PP(d_T = 0) \\ &\leq \pi w_0 \alpha + (1 - \pi) w_1 \beta = \mathcal{R}_s(d_{M_c^*}), \end{aligned} \quad (4.19)$$

where the inequality is due to $\delta \in \mathcal{C}_{\alpha,\beta}^P$ and the second equality follows from the assumption that $P_0(d_{M_c^*} = 1) = \alpha$ and $PP(d_{M_c^*} = 0) = \beta$.

From (3.11)–(3.12) and the definition of A_c and B_c in (4.16) we have

$$\begin{aligned} \mathcal{R}_s(d_{M_c^*}) &= \pi w_0 P_0(d_{M_c^*} = 1) + (1 - \pi) w_1 PP(d_{M_c^*} = 0) \\ &\leq \pi w_0 \frac{|\mathbf{q}_1|}{B_c} + (1 - \pi) w_1 \sum_{i=1}^K p_i \frac{1}{A_c q_0^i} \\ &\leq |\mathbf{q}_1| (1 - \pi) c + \sum_{i=1}^K p_i \frac{\pi c}{q_0^i} \leq (Q - 1) c, \end{aligned} \quad (4.20)$$

where $Q > 1$ is some constant that does not depend on c or π .

Fix $\epsilon > 0$ and introduce the following sequential test

$$T_{\epsilon c} = \min\{M_{\epsilon c}^*, T\}, \quad d_{T_{\epsilon c}} = d_T \mathbb{1}_{\{T \leq M_{\epsilon c}^*\}} + d_{M_{\epsilon c}^*} \mathbb{1}_{\{T > M_{\epsilon c}^*\}}.$$

Obviously,

$$\begin{aligned} \mathcal{R}_s(d_{T_{\epsilon c}}) &\leq \mathcal{R}_s(d_T) + \mathcal{R}_s(d_{M_{\epsilon c}^*}) \leq \mathcal{R}_s(d_{M_c^*}) + \mathcal{R}_s(d_{M_{\epsilon c}^*}) \\ &\leq \mathcal{R}_s(d_{M_c^*}) + (Q - 1) c \epsilon, \end{aligned} \quad (4.21)$$

where the first inequality is due to (4.19) and the second one is due to (4.20).

Since M_c^* is almost Bayes (recall (4.17)), for all sufficiently small c

$$\mathcal{R}_c(M_c^*) + \mathcal{R}_s(d_{M_c^*}) \leq \mathcal{R}_c(T_{\epsilon c}) + \mathcal{R}_s(d_{T_{\epsilon c}}) + c \epsilon. \quad (4.22)$$

Then, from (4.21) we obtain $\mathcal{R}_c(M_c^*) \leq \mathcal{R}_c(T_{ec}) + Q c \epsilon$, and consequently,

$$\begin{aligned} \pi \mathbb{E}_0[M_c^*] + (1 - \pi) \mathbb{E}^P[M_c^*] &\leq \pi \mathbb{E}_0[T_{ec}] + (1 - \pi) \mathbb{E}^P[T_{ec}] + Q \epsilon \\ &\leq \pi \mathbb{E}_0[M_{ec}^*] + (1 - \pi) \mathbb{E}^P[T] + Q \epsilon, \end{aligned} \quad (4.23)$$

where the second inequality follows from the definition of T_{ec} . Rearranging terms, we obtain from (4.23) that

$$\mathbb{E}^P[M_c^*] - \mathbb{E}^P[T] \leq \frac{\pi}{1 - \pi} \left(\mathbb{E}_0[M_{ec}^*] - \mathbb{E}_0[M_c^*] \right) + \frac{Q \epsilon}{1 - \pi}. \quad (4.24)$$

Since the last inequality holds for any $\pi \in (0, 1)$, we can set $\pi = \epsilon/(1 + \epsilon)$, which implies $B_c = \epsilon w_0/c$ and $A_c = w_1/(\epsilon c)$, whereas (4.24) becomes

$$\mathbb{E}^P[M_c^*] - \mathbb{E}^P[T] \leq \epsilon (\mathbb{E}_0[M_{ec}^*] - \mathbb{E}_0[M_c^*]) + Q \epsilon(1 + \epsilon). \quad (4.25)$$

But from (3.28) and (3.31) it follows that as $c \rightarrow 0$

$$I_0 (\mathbb{E}_0[M_{ec}^*] - \mathbb{E}_0[M_c^*]) = O(\log A_{ec} - \log A_c)$$

and from (4.16) we have $\log A_{ec} - \log A_c = |\log \epsilon| + O(1)$ as $c \rightarrow 0$, which completes the proof. \square

Remark 2. With a similar argument as the one used in the proof of Theorem 4 it can be shown that if $\mathbb{P}_0(d_{M^*(\mathbf{p})} = 1) = \alpha$ and $\mathbb{P}^P(d_{M^*(\mathbf{p})} = 0) = \beta$, then

$$\inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \mathbb{E}_0[T] \geq \inf_{\delta \in \mathcal{C}_{\alpha, \beta}^P} \mathbb{E}_0[T] = \mathbb{E}_0[M^*(\mathbf{p})] + o(1)$$

and similarly for δ_{gl} . However, the right-hand side in this asymptotic lower bound is generally not attained by $\delta_{\text{mi}}^*(\mathbf{p})$ or $\delta_{\text{gl}}^*(\mathbf{p})$ when their thresholds are selected so that $\delta_{\text{mi}}, \delta_{\text{gl}} \in \mathcal{C}_{\alpha, \beta}$.

Remark 3. While we have no rigorous proof, we strongly believe that the assertions of Theorem 4 (as well as of Theorem 5 below) hold true in the more general case where α and β approach zero in such a way that the ratio $\log \alpha / \log \beta$ is bounded away from zero and infinity, which allows one to cover the asymptotically asymmetric case as well.

4.3. Almost minimaxity

For any stopping time T and $1 \leq i \leq K$, we set $\mathcal{I}_i[T] = I_i \mathbb{E}_i[T]$. Without loss of generality, we restrict ourselves to \mathbb{P}_i -integrable stopping times, thus, from Wald's identity it follows that

$$\mathcal{I}_i[T] = \mathbb{E}_i[Z_T^i] = \mathbb{E}_i \left[\log \frac{d\mathbb{P}_i}{d\mathbb{P}_0} \Big|_{\mathcal{F}_T} \right].$$

In other words, $\mathcal{I}_i[T]$ is the expected KL divergence between P_i and P_0 that is accumulated up to time T . Let $\hat{\mathbf{p}} = (\hat{p}_1, \dots, \hat{p}_K)$ denote the prior distribution for which

$$\hat{p}_i = \frac{\mathcal{L}_i e^{\kappa_i}}{\sum_{j=1}^K \mathcal{L}_j e^{\kappa_j}}, \quad 1 \leq i \leq K. \quad (4.26)$$

Then, from (3.26)–(3.27) it follows that $\hat{\mathbf{p}}$ (almost) equalizes the KL-divergence that is accumulated by both the MiLRT and the WGLRT until stopping, in the sense that $\mathcal{I}_i[M^*(\hat{\mathbf{p}})]$ and $\mathcal{I}_i[N^*(\hat{\mathbf{p}})]$ are independent of i up to an $o(1)$ term. Indeed,

$$\mathcal{I}_i[M^*(\hat{\mathbf{p}})] = \log B + \log \left(\sum_{j=1}^K e^{\mathcal{L}_j \kappa_j} \right) + o(1), \quad (4.27)$$

$$\mathcal{I}_i[N^*(\hat{\mathbf{p}})] = \log B + \log \left(\sum_{j=1}^K e^{\mathcal{L}_j \kappa_j} \right) + o(1), \quad (4.28)$$

where only negligible terms $o(1)$ may depend on i . If additionally B is selected so that $P(d_{M^*(\hat{\mathbf{p}})} = 1) \sim \alpha$ and $P(d_{N^*(\hat{\mathbf{p}})} = 1) \sim \alpha$, then (3.33)–(3.34) imply that for every $1 \leq i \leq K$,

$$\mathcal{I}_i[M^*(\hat{\mathbf{p}})] = |\log \alpha| + \log \left(\sum_{j=1}^K \gamma_j e^{\kappa_j} \right) + o(1), \quad (4.29)$$

$$\mathcal{I}_i[N^*(\hat{\mathbf{p}})] \leq |\log \alpha| + \log \left(\sum_{j=1}^K \gamma_j e^{\kappa_j} \right) + o(1), \quad (4.30)$$

and consequently, if we denote by $\hat{\mathcal{I}}[T] = \max_{1 \leq i \leq K} \mathcal{I}_i[T]$ the maximal expected KL-divergence until stopping, we have

$$\hat{\mathcal{I}}[M^*(\hat{\mathbf{p}})] = |\log \alpha| + \log \left(\sum_{j=1}^K \gamma_j e^{\kappa_j} \right) + o(1), \quad (4.31)$$

$$\hat{\mathcal{I}}[N^*(\hat{\mathbf{p}})] \leq |\log \alpha| + \log \left(\sum_{j=1}^K \gamma_j e^{\kappa_j} \right) + o(1). \quad (4.32)$$

The following theorem states that $\delta_{\text{mi}}(\hat{\mathbf{p}})$ and $\delta_{\text{gl}}(\hat{\mathbf{p}})$ are almost minimax in this KL-sense.

Theorem 5. *Suppose that conditions (A1)–(A3) hold with $k = 1$, i.e., $\alpha, \beta \rightarrow 0$ so that $|\log \alpha| \sim |\log \beta|$. Then,*

$$\inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \hat{\mathcal{I}}[T] = |\log \alpha| + \log \left(\sum_{j=1}^K \gamma_j e^{\kappa_j} \right) + o(1). \quad (4.33)$$

If additionally A, B are selected so that $\delta_{\text{mi}}(\hat{\mathbf{p}}), \delta_{\text{gl}}(\hat{\mathbf{p}}) \in \mathcal{C}_{\alpha, \beta}$ and $k_0 = 1$, i.e., $P(d_{M^*}(\hat{\mathbf{p}}) = 1) \sim \alpha$ and $P(d_{N^*}(\hat{\mathbf{p}}) = 1) \sim \alpha$, then

$$\inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \hat{\mathcal{I}}[T] = \hat{\mathcal{I}}[M^*(\hat{\mathbf{p}})] + o(1), \quad (4.34)$$

$$\inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \hat{\mathcal{I}}[T] = \hat{\mathcal{I}}[N^*(\hat{\mathbf{p}})] + o(1). \quad (4.35)$$

Proof. Suppose that thresholds A and B are selected so that $\delta_{\text{mi}}(\hat{\mathbf{p}}) \in \mathcal{C}_{\alpha, \beta}$ and $P(d_{M^*}(\hat{\mathbf{p}}) = 1) \sim \alpha$. From Theorem 4 it follows that

$$\begin{aligned} \sum_{i=1}^K \hat{p}_i E_i[M^*(\hat{\mathbf{p}})] + o(1) &\leq \inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \sum_{i=1}^K \hat{p}_i E_i[T] = \inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \sum_{i=1}^K \frac{\hat{p}_i}{I_i} \mathcal{I}_i[T] \\ &\leq \left(\sum_{i=1}^K \frac{\hat{p}_i}{I_i} \right) \inf_{\delta \in \mathcal{C}_{\alpha, \beta}} \hat{\mathcal{I}}[T], \end{aligned} \quad (4.36)$$

whereas from (4.29) and (4.31) we have

$$\sum_{i=1}^K \hat{p}_i E_i[M^*(\hat{\mathbf{p}})] = \left(\sum_{i=1}^K \frac{\hat{p}_i}{I_i} \right) \left[|\log \alpha| + \log \left(\sum_{j=1}^K \gamma_j e^{\kappa_j} \right) + o(1) \right] \quad (4.37)$$

$$= \left(\sum_{i=1}^K \frac{\hat{p}_i}{I_i} \right) \hat{\mathcal{I}}[M^*(\hat{\mathbf{p}})]. \quad (4.38)$$

From (4.36) and (4.37) we obtain (4.33), whereas from (4.36) and (4.38) we obtain (4.34). Finally, from (4.32) and (4.33) we obtain (4.35). \square

5. How to Select \mathbf{p} ?

In this section, we consider the specification of the prior distribution \mathbf{p} , which determines the weights \mathbf{q}_0 and \mathbf{q}_1 of the MiLRT and the WGLRT when the weights are selected according to (1.6). Our goal is to select a *robust* prior, which inflicts a small performance loss under every scenario. In other words, we want to avoid a prior distribution that leads to sequential tests with very good behavior for some densities in H_1 , but with poor behavior for others.

5.1. Performance measures

We will quantify the “performance loss” of the MiLRT (and similarly for the WGLRT) under P_i by the following measure,

$$\mathcal{J}_i(\mathbf{p}) = \frac{E_i[M^*(\mathbf{p})] - E_i[S^i]}{E_i[S^i]}, \quad 1 \leq i \leq K,$$

where we recall that S^i is the SPRT for testing f_0 against f_i . That is, $\mathcal{J}_i(\mathbf{p})$ represents the *additional* expected sample size due to the uncertainty in the alternative hypothesis divided by the smallest possible expected sample size that is required for testing f_0 against f_i . Moreover, if S^i has error probabilities α and β , assumptions (A1)–(A3) hold and $k_0 = 1$, then from (2.5) and (4.12) it follows that

$$\mathcal{J}_i(\mathbf{p}) \approx \frac{C_i(\mathbf{p})}{|\log \alpha| + \kappa_i + \log \gamma_i} = \frac{\log \left[\sum_{j=1}^K (p_j / I_j) \right] + \log I_i - \log p_i}{|\log \alpha| + \kappa_i + \log \gamma_i}, \quad (5.1)$$

where by \approx we mean that the two sides differ by an $o(1)$ term. From this expression we can see that the magnitude of $\mathcal{J}_i(\mathbf{p})$ is mainly determined by K , the cardinality of \mathcal{A}_1 , and the probability of type-I error α . In particular, for every $1 \leq i \leq K$ and \mathbf{p} , $\mathcal{J}_i(\mathbf{p})$ will be “small” when $|\log \alpha|$ is much larger than $\log K$, which implies that the choice of \mathbf{p} may make a difference only when $|\log \alpha|$ is not much larger than $\log K$.

TABLE 1. Asymptotic performance loss for different prior distributions

p_i	q_1^i	$C_i(\mathbf{p})$
\mathcal{L}_i	1	$-\log(\gamma_i) + \log\left(\sum_{j=1}^K \gamma_j\right)$
I_i	$1/\gamma_i$	$\log K$
$e^{\kappa_i} \mathcal{L}_i$	e^{κ_i}	$-\log(\gamma_i e^{\kappa_i}) + \log\left(\sum_{j=1}^K \gamma_j e^{\kappa_j}\right)$
1	$1/\mathcal{L}_i$	$\log(I_i) + \log\left(\sum_{j=1}^K (1/I_j)\right)$

Moreover, from (5.1) it is clear that a good choice for \mathbf{p} would guarantee that $C_i(\mathbf{p})$ is “small” for every $1 \leq i \leq K$. In Table 1, we present $C_i(\mathbf{p})$ for the almost least favorable distribution $\hat{\mathbf{p}}$, defined in (4.26), as well as for some other intuitively appealing choices of \mathbf{p} . In particular, we consider the priors \mathbf{p}^I , $\mathbf{p}^{\mathcal{L}}$, \mathbf{p}^u which are defined so that

$$p_i^I \propto I_i, \quad p_i^{\mathcal{L}} \propto \mathcal{L}_i, \quad p_i^u \propto 1, \quad 1 \leq i \leq K.$$

Note that $\mathbf{p}^{\mathcal{L}}$, \mathbf{p}^I , $\hat{\mathbf{p}}$ are ranked, in the sense that $\mathcal{L}_i \leq I_i \leq e^{\kappa_i} \mathcal{L}_i$, since $\mathcal{L}_i = \gamma_i I_i$ and $\gamma_i \leq 1 \leq e^{\kappa_i} \gamma_i$. Thus, $\mathbf{p}^{\mathcal{L}}$ (resp. $\hat{\mathbf{p}}$) assigns relatively less (resp. more) weight than \mathbf{p}^I to a hypothesis as its “signal-to-noise ratio” increases. Note also that $\mathbf{p}^{\mathcal{L}}$ and $\hat{\mathbf{p}}$ reduce to \mathbf{p}^I when there is no overshoot effect, in which case $\kappa_i = 0$ and $\gamma_i = 1$,

whereas all these three priors reduce to \mathbf{p}^u in the symmetric case where I_i and \mathcal{H}_i do not depend on i .

5.2. Numerical comparisons

In order to make some concrete comparisons, we focus on the multichannel setup (1.5), assuming that $\{g_0^i, g_1^i\}$ can be embedded in a parametric family $g(x; \theta)$, so that

$$g_0^i(x) = g(x; \theta = 0) \quad \text{and} \quad g_1^i(x) = g(x; \theta_i), \quad 1 \leq i \leq K, \quad (5.2)$$

where $\theta_i > 0$ expresses the “signal-to-noise ratio” in channel i , $1 \leq i \leq K$.

Consider the exponential model assuming that

$$g(x; \theta) = \frac{1}{1 + \theta} e^{-x/(1+\theta)}, \quad x > 0. \quad (5.3)$$

Then I_i , κ_i and γ_i take the following form

$$I_i = \theta_i - \log(1 + \theta_i), \quad \kappa_i = \theta_i, \quad \gamma_i = (1 + \theta_i)^{-1}.$$

For the Gaussian model $g(x) = \mathcal{N}(x; \theta, 1)$, where $\mathcal{N}(x; \mu, \sigma)$ is density of the normal distribution with mean μ and standard deviation σ , the above quantities become

$$\begin{aligned} I_i &= \frac{\theta_i^2}{2}, \quad \gamma_i = \frac{1}{I_i} \exp \left\{ -2 \sum_{n=1}^{\infty} \frac{1}{n} \Phi \left(-\frac{\theta_i}{2} \sqrt{n} \right) \right\}, \\ \kappa_i &= 1 + \frac{\theta_i^2}{4} - i \sum_{n=1}^{\infty} \left[\frac{1}{\sqrt{n}} \phi \left(\frac{\theta_i}{2} \sqrt{n} \right) - \frac{\theta_i}{2} \Phi \left(-\frac{\theta_i}{2} \sqrt{n} \right) \right]. \end{aligned}$$

Assume, for simplicity, that $\theta_i = 4$ for $1 \leq i \leq K/2$ and $\theta_i = \theta$ for $K/2 < i \leq K$. Thus, the “expected” signal in the first (resp. last) channel is stronger (resp. weaker) than the signal in the last (resp. first) channel when $\theta < 4$ (resp. $\theta > 4$).

Our goal is to evaluate $\mathcal{J}_1(\mathbf{p})$ and $\mathcal{J}_K(\mathbf{p})$, i.e., the inflicted performance loss when signal is present in the first and last channel respectively, as a function of θ , for different prior distributions. We do so using asymptotic approximation (5.1), in which we have set $K = 10$ and $\alpha = 10^{-4}$, and we present the results for the exponential case in Figure 1 and for the Gaussian case in Figure 2.

The plots in both figures show that setting $\mathbf{p} = \hat{\mathbf{p}}$ (resp. $\mathbf{p} = \mathbf{p}^u$) leads to a better performance when signal is present in the channel with stronger (resp. weaker) signal-to-noise ratio. However, the inflicted performance loss when the signal is present in the other channel can be very high. On the other hand, setting $\mathbf{p} = \mathbf{p}^I$ or $\mathbf{p} = \mathbf{p}^C$ leads to a more robust performance, since the performance loss is similar (and relatively

small) irrespectively of the channel in which signal is present and of the relative signal strengths.

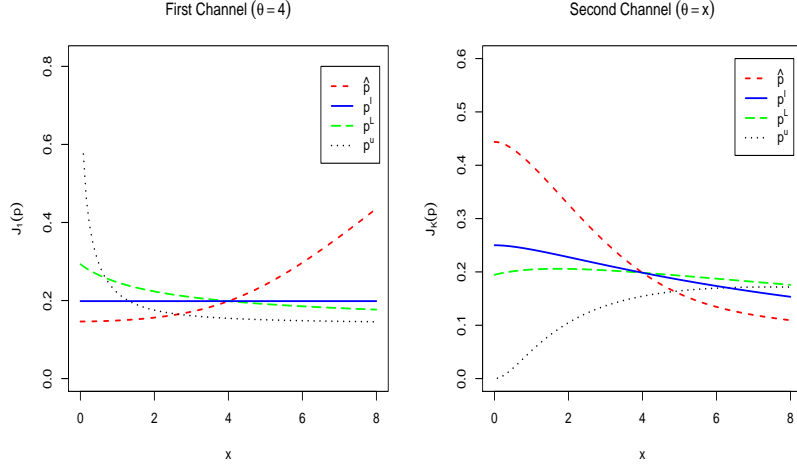


FIGURE 1. Performance loss for different prior distributions in a multichannel problem with exponential data.

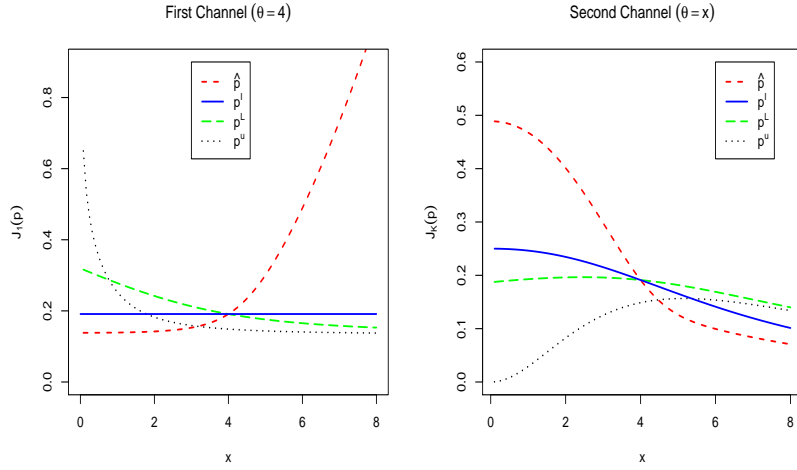


FIGURE 2. Performance loss for different prior distributions in a multichannel problem with Gaussian data.

6. Monte Carlo Simulations

In this section, we present a simulation study whose goal is to check the accuracy of the asymptotic approximations established in Section 4 and to compare the MiLRT with the WGLRT for realistic probabilities of errors. In particular, we consider the multichannel setup (1.5) with $K = 3$ channels, exponential distributions given by (5.2)–(5.3) and parameter values selected according to Table 2. Since our main emphasis is on the fast detection of signal, we set $\beta = 10^{-2}$ and consider different values of α . Moreover, we choose the thresholds A and B according to (3.21), whereas we select the weights according to (1.6) with $\mathbf{p} = \mathbf{p}^{\mathcal{I}}$.

TABLE 2. Parameter values in a multichannel problem with exponential data

θ_i	I_i	κ_i	γ_i	q_1^i	q_0^i
0.5	0.095	0.5	0.67	0.308	0.013
1	0.584	1	0.4	0.837	0.078
2	0.901	2	0.33	1.380	0.138

In the first three columns of Table 3 we compare the type-I error probabilities for the two tests, which have been computed based on simulation experiments, against the target level α . More specifically, these error probabilities are computed using representations (3.16) and (3.18) and *importance sampling*, a simulation technique whose application in Sequential Analysis goes back to Siegmund (1976). These results indicate that selecting B according to (3.20) leads to type-I error probabilities very close to α for both tests, even for relatively large α . In particular, we see that $P_0(d_{M^*} = 1)$ is slightly larger than α , which is expected, since (3.20) implies $P_0(d_{M^*} = 1) \sim \alpha$, whereas we also observe that α is a sharp upper bound for $P_0(d_{N^*} = 1)$, the type-I error probability of the WGLRT.

In the remaining columns of Table 3, we present for both tests the (simulated) expected sample size under P_i , $i = 1, 2, 3$ and in Figure 6 we plot these values against the corresponding (simulated) type-I error probabilities. In these graphs, we also superimpose asymptotic approximation (3.33) (dashed line), as well as the asymptotic performance of the corresponding SPRT, (2.5), which is given by the solid line. Triangles correspond to the WGLRT and circles to the MiLRT. From these results we can see, first

TABLE 3. Type-I error probabilities and the expected sample sizes under P_i , $i = 1, 2, 3$ for different values of the target probability α when $\beta = 10^{-2}$.

α	$\frac{P_0(d_{M^*}=1)}{\alpha}$	$\frac{P_0(d_{N^*}=1)}{\alpha}$	$E_1[M^*]$	$E_1[N^*]$	$E_2[M^*]$	$E_2[N^*]$	$E_3[M^*]$	$E_3[N^*]$
10^{-2}	1.051	0.994	59.9	59.4	17.8	19.4	6.2	7.3
10^{-3}	1.033	0.995	84.1	84.1	25.7	27.1	9.0	9.9
10^{-4}	1.025	0.996	108.5	108.3	33.7	34.6	11.7	12.4
10^{-5}	1.017	0.996	132.5	132.3	41.4	42.0	14.3	15.0

of all, that asymptotic approximation (3.33) is very accurate for both tests. Moreover, we can see that the two tests have similar performance. In particular, their performance is identical when signal is present in the channel with the smallest signal strength. In the other two cases, the MiLRT seems to perform slightly better, however the difference is small.

7. Conclusion

In this work, we performed a detailed analysis and optimization of weighted GLR and mixture-based sequential tests when the null hypothesis is simple and the alternative hypothesis is composite but discrete. Irrespectively of the choice of weights, both tests minimize asymptotically, at least to first order and often to second order, the expected sample size under each possible scenario as error probabilities go to 0. However, with appropriate selection of weights, both test achieve higher-order asymptotic optimality properties. Specifically, they minimize a weighted expected sample size as well as the expected Kullback–Leibler divergence in the least favorable scenario to within asymptotically negligible terms as error probabilities go to zero. Moreover, based on simulation experiments, we can conclude that the two tests perform similarly even for not too small error probabilities. Finally, we believe that the proposed approach can be extended to sequential testing of multiple hypotheses, a substantially more complex problem that we plan to consider elsewhere.

Acknowledgements

We are grateful to a referee whose comments improved the presentation. This work was supported by the U.S. Air Force Office of Scientific Research under MURI grant

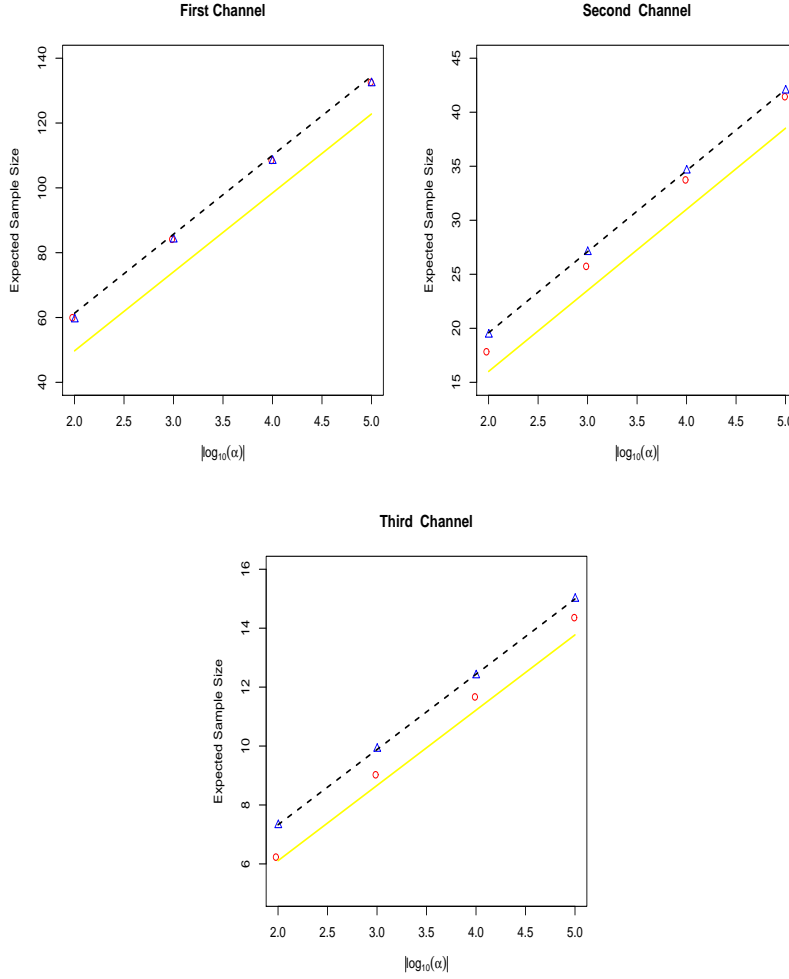


FIGURE 3. Expected sample size of MiLRT and WGLRT under P_i against type-I error probability (in logarithmic scale), $i = 1, 2, 3$. The dashed line represents asymptotic approximation (3.33), whereas the solid line refers to (2.5), the asymptotic performance of the corresponding SPRT. The triangles (resp. circles) represent the simulated performance of the WGLRT (resp. MiLRT).

FA9550-10-1-0569, by the U.S. Defense Threat Reduction Agency under grant HDTRA1-10-1-0086, by the U.S. Defense Advanced Research Projects Agency under grant W911NF-12-1-0034 and by the U.S. National Science Foundation under grants CCF-0830419,

EFRI-1025043, and DMS-1221888 at the University of Southern California, Department of Mathematics.

References

- Chernoff, H. (1972). *Sequential Analysis and Optimal Design*. SIAM, Philadelphia.
- Chow, Y. S. , Robbins, H. and Siegmund, D. (1971). *Great Expectations: The Theory of Optimal Stopping*. Houghton Mifflin, Boston.
- Bhattacharya, R. N. and Rao, R. R. (1986). *Normal Approximation and Asymptotic Expansions*. Wiley, New York.
- Dragalin, V. (1999). Asymptotics for a sequential selection procedure. *Statist. Decisions* **4**, 123-137.
- Dragalin, V. P., Tartakovsky, A. G., and Veeravalli, V. V. (2000). Multihypothesis sequential probability ratio tests - Part II: Accurate asymptotic expansions for the expected sample size. *IEEE Trans. Inform. Theory* **46**, 1366-1343.
- Dragalin, V. and Novikov, A. (1999). Adaptive sequential tests for composite hypotheses. *Surveys in Applied and Industrial Mathematics, TVP press* **6**, 387-398.
- Fellouris, G. and Tartakovsky, A. G. (2012). Nearly minimax mixture-based open-ended sequential tests. *Sequential Analysis* **31**, 297-325.
- Gut, A. (2008). *Stopped Random Walks, Second Edition*, Springer.
- Kiefer, J. and Sacks, J. (1963). Asymptotically optimal sequential inference and design, *Ann. Math. Statist.* **34**, 705-750.
- Lai, T. L. (1988). Nearly optimal sequential tests of composite hypotheses. *Ann. Statist.* **16**, 856-886.
- Lai, T. L. (2004). Likelihood ratio identities and their applications to sequential analysis. *Sequential Analysis* **23**, 467-497.
- Lai, T. L. and Siegmund, D. (1977). A nonlinear renewal theory with applications to sequential analysis I. *Ann. Statist.* **5**, 628-643.
- Lai, T. L. and Siegmund, D. (1979). A nonlinear renewal theory with applications to sequential analysis II. *Ann. Statist.* **7**, 60-76.
- Lorden, G. (1967). Integrated risk of asymptotically Bayes sequential tests. *Ann. Math. Statist.* **38**, 1399-1422.
- Lorden, G. (1973). Open-ended tests for Koopman–Darmois families. *Ann. Statist.* **1**, 633-643.

- Lorden, G. (1977). Nearly optimal sequential tests for finitely many parameter values. *Ann. Statist.* **5**, 1-21.
- Pavlov, I. V. (1990). A sequential procedure for testing composite hypotheses with application to the Kiefer–Weiss problem. *Theory Probab. Appl.* **35**, 280-292.
- Pollak, M. and Siegmund, D. (1975). Approximations to the expected sample size of certain sequential tests. *Ann. Statist.* **3**, 1267-1282.
- Robbins, H. and Siegmund, D. (1970). A class of stopping rules for testing parameter hypotheses. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* **4** (Edited by K. M. Le Cam, J. Neyman, and E. L. Scott), 37-41. University of California Press, Berkeley.
- Robbins, H. and Siegmund, D. (1974). The expected sample size of some tests of power one. *Ann. Math. Statist.* **2**, 415-436.
- Schwarz, G. (1962). Asymptotic shapes of Bayes sequential testing regions. *Ann. Math. Statist.* **33**, 224-236.
- Siegmund, D. (1975). Error probabilities and average sample number of the sequential probability ratio test. *J. Roy. Statist. Soc. Ser. B* **37**, 394-401.
- Siegmund, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. *Ann. Statist.* **4**, 673-684.
- Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer-Verlag, New York.
- Tartakovsky, A. G. (1991). *Sequential Methods in the Theory of Information Systems*. Radio & Communications, Moscow.
- Tartakovsky, A. G., Li, X. R., and Yarovoy, G. (2003). Sequential detection of targets in multichannel systems. *IEEE Trans. Inform. Theory* **49**, 425-445.
- Tartakovsky, A. G., Rozovskii, B. K., Blažek, R. B., and Kim, H. (2006). Detection of intrusions in information systems by sequential change-point methods. *Statistical Methodol.* **3**, 252-293.
- Wald, A. (1944). Sequential tests of statistical hypotheses. *Ann. Math. Statist.* **16**, 117-186.
- Wald, A. and Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Statist.* **19**, 326-339.
- Woodroffe, M. (1976). A renewal theorem for curved boundaries and moments of first passage times. *Ann. Probab.* **4**, 67-80.

Woodroffe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. SIAM, Philadelphia.

Zhang, C.H. (1988). A nonlinear renewal theory. *Ann. Probab.* **16**, 793-825.

University of Southern California, Department of Mathematics, 3620 S. Vermont Avenue, Los Angeles, CA 90089-2532, USA

E-mail: fellouri@usc.edu

University of Southern California, Department of Mathematics, 3620 S. Vermont Avenue, Los Angeles, CA 90089-2532, USA

E-mail: tartakov@usc.edu

(Received March 2012; Revised January 2013; Accepted)